



The National Center on
Education in the Inner Cities

The CEIC REVIEW

A catalyst for merging research, policy, and practice.

Volume 10 • Number 7 • September 2001

Making Schools Work

Recommendations from a National Invitational Conference on Improving Educational Productivity: Lessons from Economics

David H. Monk, Pennsylvania State University; Herbert J. Walberg, University of Chicago at Illinois; and Margaret C. Wang, Temple University

Educational researchers and policymakers, traditionally concerned with promoting access to opportunity and equity, have in the past decade shown growing interest in productivity, motivated by the increasing concern for accountability and efficiency in the nation's school systems. Building on earlier work investigating the relation between educational inputs and outcomes, a new generation of economists is applying insights and techniques to further our understanding of what makes schools more productive and efficient.

Economists and Educators Collaborating

To bring together these new economists of education, showcase their work, and facilitate exploration of its implications for practice, a national invitational conference was conceived. Cosponsored by the Laboratory for Student Success and the National Center on Education in the Inner Cities, both at Temple University's Center for Research in Human Development and Education, this conference on "Improving Educational Productivity: Lessons from Economics" was held November 27-28, 2000 in Washington, DC.

An important goal of the conference was bridging the divide that can exist between the worlds of econom-

ics and educational policy. An economic understanding of the impact of resources on student success is important to school efficiency, yet many educators remain wary of economists as outsiders with limited knowledge of schools. Thus the scholars who contributed commissioned papers to the conference were encouraged to think of policymakers at all levels as the intended audience for their reports. To help policymakers in their decisions, the papers were intended to explain and interpret the latest, most promising research on educational productivity and efficiency.

These papers (summarized here by the editors and recently published by Information Age Publishing) presented overviews, critiques, and findings with rich implications for strengthening our schools. They approached the overarching issues of productivity and efficiency in various ways. The first three papers focused on a series of external influences on the operation of public schools, including changes in taxpayer support of education (Downes and Figlio), market pressures from vouchers and private schools (Goldhaber), and alternative funding distribution mechanisms in response to litigation (Murray). The next four papers concerned relationships between pro-

ductivity and reforms in schooling practice. Productivity research on grade retention (Eide), teacher quality (Loeb), school-level efficiency (Schwartz and Stiefel), and internal resource-allocation practices (Rubenstein and Iatarola) was analyzed with an emphasis on implications for policy. The final set of papers reported on insights from new approaches to the study of educational productivity, including the need to consider individual differences among students (Taylor), the difficulties resulting from looking exclusively at average achievement gains (Hussain), and the estimation of school effects using sophisticated econometric tools (Ludwig). The conference participants included practitioners, policymakers, and researchers from multiple disciplines. The conference design facilitated a dialogue among these groups and gave rise to recommendations for improved research and applications to policy and practice.

Next-Step Recommendations

The participants also worked in small groups with the conference leaders to develop recommendations for next steps towards improving educational productivity. Recommendations were

(**Making**, continued on p. 22)



The National Center on Education in the Inner Cities is a unit in the Temple University Center for Research in Human Development and Education, an interdisciplinary center devoted to fostering healthy developmental and educational success of children and families in this nation's urban communities. Inquiries about the work of the Center should be sent to Information Services, CRHDE, Temple University, 1301 Cecil B. Moore Avenue, Philadelphia, PA 19122-6091. Copyright © 2001

Tax Revolts and School Performance

Thomas Downes, Tufts University; and David Figlio, University of Florida and National Bureau of Economic Research

The 1978 passage of Proposition 13 limiting property taxes in California stimulated a tax revolt in America that was meant to reduce waste in government services, including public schools. Many states and localities subsequently passed both tax and government spending limits in a fiscal-reform movement that peaked in the early 1980s but continues today. Although these restrictions were of heterogeneous scope and severity, the similarity of effects across states has seemed more important than have differences from state to state, as researchers over the past 20 years reached consensus that the limits significantly reduced public services in the affected areas.

However, as proponents have argued, reduction in service quantity and quality does not necessarily follow from reduction in revenue or spending. Service can improve after limits are imposed if waste is reduced and if resources are used more efficiently. For schools in particular, there seems no reason to believe that reduced spending would affect the quality of student performance, since research has shown no effect on student performance with increased spending.

Nevertheless, because research has also shown that tax limits can have extrafinancial effects, such as reduction in community homogeneity, that may indirectly influence public services like schools, economic models that posit simple causal relations between fiscal limits and public-sector changes seem unsatisfactory. In light of investigation of the complex effects of fiscal limits on public-school student performance, this study focuses on that public-sector outcome and proposes a series of potential explanations for the relation found in the literature between fiscal limits and negative student performance. The authors draw analogies between the

tax revolt of the recent past and the current accountability movement.

Effects of Tax and Spending Limits on Student Performance

Despite the long delay between implementation of the major statewide tax and spending restrictions and investigation of their influence on student performance, early research on the limits' general effects indicated reductions in the quality of school services, like California's cuts in service employees and in the diversity of course offerings. The first research on the performance effects of tax limits, conducted in 1992 in California, showed no changes in performance on achievement tests, but since the state also imposed relevant financial reforms, no general conclusions could be drawn from this study.

More recent research has examined nationwide effects, using a cross section of student data from the National Educational Longitudinal Survey (NELS). It was found that the scores on achievement tests of 10th graders were significantly lower in states faced with either revenue or expenditure limits, but the study could not rule out the possibility that sorting and unobserved tastes for education caused the test-score differences and the passage of limits. Another investigation of Chicago suburban districts faced with property-tax restrictions eliminated that problem and showed modest performance effects: slowed growth in the performance of third graders on achievement tests in mathematics. This work was limited by the possible exceptionality of the districts studied and by the lack of long-term data.

Building on insight from this research, the present study's authors recently examined effects of limits on performance more longitudinally, comparing public-school performance data from the National Longitudinal

Study of the High School Class of 1972 and from the 1992 (senior-year) follow-up of the NELS. This comparison allowed long-term analysis of limits' effects, and further comparison of states with limits to those without afforded a control group. The authors' analysis found that local fiscal limits correlated significantly with reduced performance on standardized tests of math skills. Reductions were from 1% to 7%, depending on model specification. However, no evidence of reduced reading-skill performance was found.

Most research on the effects of fiscal limits has focused on students who remain in public schools. However, recent work has shown that the limits affect enrollment, with public schools' share of the potential student population dropping where the most initial cuts are needed after a tax restriction as a result of both dropouts and private-school enrollment shifts. Thus research should also focus on the effects of limits on private-school performance. Using methodology similar to their public-school study, the present authors have found modest evidence of negative effects of tax limits on private-school test scores. It may be that those curtailments lower the quality of private-school achievement because of lower public-school competition, higher private-school selection by less capable students after the imposition of limits, and related effects of lower peer competition.

On balance, evidence on the impact of limits on performance is more compelling for mean performance than for cross-district distribution. Performance does deteriorate more rapidly in economically disadvantaged areas, though frequently not at statistically significant levels. While further investigation is needed into the dependence of the effects of limits on conditions in particular districts,

research increasingly suggests that tax and spending limits lower academic achievement.

Explaining Effects of Limits on Performance

Since the relationship between increased spending and improved student outcomes is weak, it seems surprising that spending limits affect outcomes. Increased spending apparently fails to affect achievement because additional resources are not used efficiently. Even allocations that are meant to help students directly, such as hiring teachers to reduce class size, show little influence. Moreover, allocations often go to administrative improvements that affect students only indirectly. Or allocations go to increase the salaries of experienced teachers, with whom performance benefits are not typically associated, not to hiring highly qualified new teachers who could be expected to improve outcomes.

Yet decreased spending does not show symmetrical impact on allocation. Whereas spending increases tend to accrue to administration, decreases tend to be borne by instruction. If instructional spending matters at all (which is still an open question), then reducing spending would more directly affect students than increasing it would. Even less funding would be available for hiring better teachers, and that would lead to changes in the instructional labor market that would be deleterious to student outcomes. Research has indeed associated tax limits with reduced average qualifications of new teachers, which in some cases (selectivity of undergraduate institution, subject-matter expertise) influence student performance negatively. Research has also shown that districts subject to tax limitations cut the salaries of starting teachers while maintaining those of experienced ones, making it likely that the schools will attract less qualified applicants.

Another cause of the negative effects of limits lies in administrative inefficiency. Neither tax nor spending

limits typically stipulate explicit incentives for eliminating administrative waste. Therefore, research shows, some administrators strategically allow high budgets and technical inefficiencies to continue. That encourages communities to override limits in response to continued declines in student performance that stem from administrative inaction or instructional cuts. Evidence that districts that are bound by limits reduce instruction but not administration supports the view that administrative waste contributes to the negative effects of limits.

Admittedly, administration is not favored over instruction in all districts. Sometimes under local limits, competitive pressure from unaffected districts causes affected ones to sustain instructional expenses and to cut administrative waste. Related research has shown that competition for resources among urban public schools tends to increase their efficiency.

Moreover, increased community heterogeneity may account for part of the negative influence of limits. Since educating a more socioeconomically and academically heterogeneous community is more costly than educating a homogenous one, fiscal constraints in such communities may increase the likelihood of reduced outcomes. Further, where fiscal limits lead some students, especially those with the highest ability and socioeconomic status, to move to the private sector, as studies suggest they do, performance levels may fall, and educating the remaining public-sector students may be even costlier.

While these causal mechanisms for the negative effects of limits on student achievement are not comprehensive, they explain why performance should drop more dramatically with fiscal reduction than it would rise with fiscal expansion. Thus the intent of voters who favor reductions in property taxes, to enjoy a financial benefit without facing a less effective public-school system, seems unreal-

ized, especially since tax reductions fail to specify how money is to be spent or to provide otherwise for eliminating inefficiencies in school systems. That the most productive component of the system, the instructional one, is the most susceptible to reduced spending exacerbates inefficiencies and increases the likelihood that the aims of fiscal limits will not be achieved.

Conclusion

Although evidence suggests that fiscal limits reduce student performance, it remains debatable whether such limits are good public policy, since more comprehensive research on their costs and benefits remains to be conducted. The tax revolt does make it clear, however, that educational reforms can have unintended consequences. If this is true for fiscal limits, it may also be true for an important current trend in school reform: increasing school-level accountability through linking student test scores to public evaluation of schools. This trend has been inspired by the notion that greater public accountability motivates schools to appear better and thus retain students.

However, since the experiments with limits show that administrators do not always allocate resources effectively, increased accountability will not necessarily lead to greater efficiency. Although resources are likely to be allocated to productive uses in an atmosphere of accountability, administrators could still siphon resources from productive uses to areas believed to improve test scores. However, these areas may not be those most conducive to increasing the long-term productivity of the schools. Thus it is not obvious from economic theory that increasing accountability increases efficiency. Just as in the case of the impact of fiscal limitations, the impact of accountability on efficiency remains ultimately an empirical question.



The Interface Between Public and Private Schooling: Market Pressure and the Impact on Performance

Dan Goldhaber, Urban Institute in the Education Policy Center

Debate about school choice and educational vouchers has centered on vouchers' effects on student outcomes. However, too often debate has overlooked the determining role of policy choices or results of voucher experiments. Also, complex school–market interaction may make outcomes impossible to predict before broad voucher implementation. This study examines theory and evidence for voucher policies, assessing their feasibility. It describes arguments underlying advocacy of school choice, research on competition in our mostly voucher-free educational system, public- and private-school differences, influences of school type on voucher experiments, and voucher policies' likely impact.

Theoretical Arguments for Choice

MARKET EFFICIENCY

Advocates believe more school choice will allow better student–school fit than public education. But others maintain better matching will not improve performance unless financial pressures improve educational practices. A voucher-based educational market linking funding to pupil counts based on parental choice could motivate improvement, especially in inefficient public schools, eliminating allocations unrelated to student achievement. This market model of school choice assumes an efficient market of equivalent products, buyers and sellers of comparable size, mobile market entry and exit, and perfect knowledge among market participants. Although no market is entirely efficient, this model's proponents believe vouchers would increase competition and yield more efficient achievement-to-cost ratios.

THE MARKET AND PUBLIC SCHOOLS

Public-school advocates argue it is inappropriate to view public schools in a market context. They note that beyond individual instructional effects, public schools influence

social cohesion, with some research suggesting more school choice would increase racial and socioeconomic segregation, thus decreasing social cohesion. But established housing patterns already relatively segregate schools today. Choice proponents maintain vouchers could break the school–housing link. This claim cannot yet be evaluated, because broad voucher systems have not yet been implemented. Another argument against vouchers has been that competition would result in fewer resources for the most needy students; however, little evidence exists on how vouchers might impact spending on public schools or distribution of spending.

Another reason to be skeptical of the market view of schools is the crucial role information plays in market efficiency. Inadequate parental information about schools' quality may result in poor choices. Many schools, though, are providing more information in public report cards. Studies show positive relations between parental choice and school quality, suggesting parents make academically beneficial decisions. This relationship seems strongly related to the school's or community's socioeconomic status. Because of unavoidable inequalities in school and student achievement, even perfect parental information and school access cannot guarantee equal distribution of gains. Whatever the market conditions, unless families base choices on academic quality, not features like proximity and cocurricular programs, increased competition may not boost achievement.

Competition's Impact on Schools

Since theory alone cannot determine vouchers' feasibility, evidence on school choice should be considered. Of many competing forms of school choice, open enrollment within and between public school districts and private–public school competition parallel most closely the likely effects

of voucher programs. Research on these choices has produced mixed results.

PUBLIC-SECTOR COMPETITION

A study of open enrollment in Chicago, where half of high-school students chose to change schools, showed changing did not significantly raise changers' graduation rates or harm those left behind, except in the case of Chicago's "Career Academies," where those attending experienced small benefits. These results suggest better school–student fit can improve outcomes. Since such intradistrict choice does not affect revenue, it may not stimulate school improvement; thus interdistrict choice, threatening student loss, seems a better model of competitive educational markets. Anecdotal evidence indicates interdistrict choice leads to innovations to attract and keep students, even when few actually move. Moreover, studies of interdistrict competition with virtual choice due to large numbers of area districts showed competition led to improved school quality and student performance.

PUBLIC–PRIVATE COMPETITION

Since an area's private and public schools compete, a market-based argument suggests competition should improve local student outcomes. Though one study found increased private-school enrollment led not to greater public-school achievement but greater resource investment, this finding seems ungeneralizable. Given public-school funding's complexities, including possible reciprocal effects of private-school enrollment on area revenues, simple correlation between private-school enrollment and public-school funding cannot be predicted. Factors like parental preference, community affluence and educational demand, and private-school establishment where public ones are weak all need accounting for. By considering

effects of public-school quality and community features on private-school supply, researchers have found competition from private schools benefits public-school students' performance. Gains are modest, with public-school test scores and graduation rates rising less than 5%, implying vouchers would not significantly elevate public-school efficiency unless students found private schools much more attractive. Differences between these school sectors need further scrutiny.

Public-Private School Differences INSTITUTIONAL DIFFERENCES

Private schools generally provide more administrative freedom and curricular rigor than bureaucratic, highly tracked public schools. Private schools also foster the effective teachers crucial to student achievement. Although public schools pay teachers more, they often hire weaker candidates and value their aptitudes less than private schools paying in greater accordance with market conditions and proven abilities. Public schools' uniform salary schedules may discourage talented candidates and fail to motivate excellence, while private schools' performance-based compensation and pay differentials may motivate achievement. Teacher characteristics like enthusiasm may account for the effectiveness of these compensation differences, but so may student characteristics. Private-school faculty may trade salary security for more advantaged pupils. Student advantages may also shape policy and practical differences between the two sectors.

STUDENT OUTCOME DIFFERENCES

Comparing test-score differences between the sectors' students has yielded mixed results. Recent evidence showed positive private-school effects, especially for urban minority pupils, but the largely nonexperimental data did not rule out alternative causes of improvement. But even if private schools' student achievement is not clearly superior, their educational cost may be. Private-school education, especially Catholic, can cost 50% less per pupil. If achievement is equivalent

and these costs are accurate, then private schools are more cost-effective. However, these estimates ignore subsidies masking additional costs. Also, private schools often enroll advantaged children, who are inherently less costly to educate. Such student differences and unobservables like parental motivation to support education hinder valid comparison of the two sectors' efficiency in producing achievement.

VOUCHER EXPERIMENTS

Since educational experiments use random assignments and control groups to eliminate the need to account for background and unobservables, they are more comprehensible and useful in gaining public support than other means. Though political constraints have limited voucher experiments, recently several providing data on school-sector differences have been privately funded, targeting low-income urban students in various grades. Students were randomly chosen by elective lottery to receive modest vouchers. Postvoucher surveys showed attending private school benefited African Americans. In two years, their test scores narrowed the national Black-White achievement gap by up to one half. These effects were large compared to other interventions like class size reduction.

However, assessments did not control for peer effects, so improvement may have been due to student, not school, quality. Also, missing effects for non-African Americans remained unexplained, and attrition may have biased results towards better pupils. Finally, imperfect randomization, possible experimental-setting influences, and failure to account for varying school quality in both sectors argue caution in generalizing from the experiments' results.

Voucher Policies' Likely Impact

While voucher experiments furnish evidence favoring general voucher implementation, the possibility of peer effects warns that a general policy rendering both school sectors demographically similar might

eliminate a key ingredient that makes private institutions motivate change. Further, nonexperimental evidence, though inconclusive, is important for predicting voucher-policy effects. The evidence of unobservables and other nonmarket explanations of achievement gain indicates optimizing these might cause improvements more efficiently than costly voucher plans. It is also unclear whether increased private-education demand would be met by better schools. Economic theory suggests that the most efficient private schools are now thriving in high-demand markets and that additional schools may be weaker than many public schools, eliminating incentives for change.

Another unknown is the voucher plans' impact on public-school funding. Again, competition might produce public education providing less social support than the current system. As voucher plans change student distribution across schools, affecting social cohesion, trade-offs are likely, as schools sacrifice qualities like diversity for achievement. Voucher programs could be designed to motivate public-school support through taxation, but the effects of such incentives are unknown, as are the effects of voucher plans' administrative costs.

Conclusions

Although results are mixed, evidence on school competition supports the notion that it improves student outcomes. Despite significant differences between public and private schools, especially in teacher compensation and student outcomes, it is unclear whether schools or students account for differences. Voucher experiments show a positive private-sector effect. But existing theory and evidence suggest future voucher policies' success depends on policy details like private-school requirements and the size, eligibility, and financing mechanism of vouchers. Though cautious optimism is warranted, socioeconomic effects of widespread voucher implementation are uncertain.



State Aid and Education Outcomes

Sheila E. Murray, RAND Corporation

Traditionally, financial control of American public schools has been local. Although states have shared in costs, their traditional major role has been to monitor curriculum, evaluation, and standards. However, since the landmark 1971 case, *Serrano v. Priest*, in California, states have assumed a larger, redistributive role in educational finance. The resulting conflicts over loss of community control have sometimes caused the withdrawal of parental and political support from public education. This report describes research on the increasing role of states in public-school finance, examining ways of achieving and evaluating funding equity as well as the impact of states' efforts on spending and on student outcomes.

Achieving Equity

EQUITY PROGRAMS

It has long been argued that local funding of public schools is unfair, since tax-base differences across districts are large. As of 1999, opponents of local funding had argued cases in state supreme courts of 43 states; educational finance reform had been mandated in 19 states. States use many programs to fund equalization efforts. The flat-grant programs typical before the 1970s, through which districts receive funds independent of district expenditures, have become less common as states have become more concerned with addressing inequality across districts. Full state education funding is also uncommon, as most localities oppose so much state control. Most common are equalization grant programs, divisible into two categories.

First, *foundation grants* are the most popular programs, used by 40 states. They are designed to guarantee that each student in a state receives a minimum of funding. Funding is calculated on a minimum founda-

tion level and a state-established minimum uniform tax rate. State aid fills the gap between resources needed according to state measures and a district's ability to meet that need. Adjustments are made to reflect cost differences across districts for student education and to allow total spending in some districts receiving aid to exceed the minimum foundation level.

Second, *district power equalization programs* (DPEs) work to equalize on the basis of local spending, not minimum need. Under DPEs, states calculate aid on the basis of the difference between a district's tax base per student and a tax-base rate established by the state. When a district's tax-base rate exceeds the state's, the state does not recover funds from the district to fund less endowed districts (and theoretically to make the program self-financing). Instead, programs are constructed so that even wealthy districts receive some state aid.

EVALUATING EQUITY PROGRAMS

The definition of equity determines the success of an equity program. If equity means equal spending throughout a state, no plan achieves equity, and only a fully state-controlled program could. If equity means that districts with the same tax rate can spend the same amount on education, DPEs, based on tax rates, seem the best means to equity. And if equity means wealth neutrality, so that communities spending more would not necessarily be the wealthiest ones, DPEs also seem to lead to equity, since they disconnect the educational resources of a district from the district's wealth. However, DPEs do not benefit all districts needing more resources. For instance, some districts with low income but with high taxable property wealth (due perhaps to an elderly population) would miss out on needed aid, as would cities

with much taxable commercial property but a high number of low-income students.

Finally, if equity means that all students achieve a minimum educational level, outcomes, not inputs, become the focus, and inequitable funding might be needed to achieve equity in districts with many disadvantaged students. This definition of equity seems to favor foundation grants, but most states have not calculated outcomes into their foundation financing. Since accounting for individual students' needs is prohibitively expensive, states must build finance systems that provide districts, schools, and teachers with resources and inducements to tailor instruction to student needs.

Statistical approaches to calculating links between financing, instructional delivery, and outcomes are being developed. One approach fixes outcomes at an agreed level and then uses student characteristics, instructional-factor prices, and statistically determined cost functions for a district to estimate the amount the district needs in order to provide adequate education. However, such approaches are flawed by their poor ability to estimate how much more spending is needed in low-income districts to achieve adequacy.

Nonstatistical approaches that attempt better links between financing and outcomes include using successful districts as benchmarks for funding, constructing delivery models, and implementing commercial whole-school designs. Yet these strategies make outcomes no more predictable, and consensus is low on what outcomes should be. Today's high-stakes testing trend, for instance, has been criticized for narrowing curricula and distorting scores.

A final approach to achieving outcome equity through efficient funding is to adjust funding for dis-

strict cost differences on the basis of measures like the consumer price index. But such adjustments fail to account for inflation or for specific community costs. In reality, most funding decisions still must rely on political bargaining, and substantive redistribution in favor of equity, however that is defined, only occurs through litigation.

The Impact on Spending

Court-ordered state finance reform has generated more equal spending across districts, mainly through directing new state funds to low-spending districts. This change has affected resource use. In the past few decades, state funding for students grew markedly. From 1972 to 1992, state contributions to public education grew from 38.3% to 49.3%. Differences between the highest and lowest student spending narrowed somewhat in that time. Moreover, at the end of the period, unequal spending between states was much larger than within them, and more than 90% of the reduction during the two decades was due to reduced inequality between states.

California, the first state to experience court-ordered reform, illustrates the impact of spending reform more particularly. Research has shown that increased state educational financing in California has led to reduction in overall spending, with about half the decline due to the 1971 *Serrano* decision. The effects on educational spending of 1978's Proposition 13, which limited property tax rates, have been hard to separate from the effects of *Serrano*, since the proposition limited local educational resources. Indeed, some have argued that *Serrano* led to Proposition 13, as taxpayers found educational returns for their taxes insufficient after *Serrano*. Nevertheless, since court-ordered educational reform led to tax limits in only three other states, this causal link seems doubtful.

Kentucky also illustrates the impact of court reform. It saw dramatic

local educational spending disparities by the 1980s between wealthier and poorer districts. State funding, which was based on the number of classrooms, not on the number of students, exacerbated this difference. In 1989, the state supreme court instructed the state to create a more equitable system. The result, the 1990 Kentucky Education Reform Act, raised the foundation grant, adjusted equalization grants and property assessments, and changed the basis for aid calculations from classrooms to students. By 1992, funding to the poorest districts had grown by over two thirds, while funding increases to wealthier districts were limited.

Further research has shown that reform litigation increases educational spending significantly, although increased state funding can decrease overall spending on education, as in California. Moreover, using econometric modeling to assess the effects of state spending reform over time, the author and her co-workers have concluded that court-mandated reforms reduced inequality within states by percentages ranging from 16% to 38%; that spending rose by 11% in the lowest spending districts, sometimes with state revenues replacing local, providing tax relief; and that increased educational spending was supported by higher taxes. Gains persisted, and benefits went consistently to those districts with lower levels of spending.

Results vary from state to state, but the author's research indicates that districts have directed about 40% of their additional resources to hiring teachers and raising their salaries. Poorer districts raised teacher numbers and salaries most. Instruction probably accounts for most of the increased expenditures, but research on funds use is hampered by a lack of detailed national data on expenditures like textbooks and computers.

The Impact on Educational Outcomes

The many studies that link educational finance reform with outcomes

have been inconclusive, showing little clear impact. Confusing the matter is the link between higher outcomes on the SAT and spending inequality, which implies that we might maximize average outcomes by spending much on a few students, those likely to excel on the SAT, and little on the rest. Another way to examine the impact of spending on performance is to examine the outcomes of court-mandated reform, which capture other causes of change, such as improvements in instruction. Research so far, using both California and national data, has shown that court-mandated reforms do not lead to significant changes in student performance.

One study attempted to account for the selectivity effects of scores on the SAT, which tends to be taken by students who come from wealthier families and who rank highly in their class. The study concluded that finance reforms modestly equalized the test scores of students of different socioeconomic backgrounds. During the 1980s, the gap in SAT scores closed by about 10 points between children of more and less educated parents in 12 court-mandated reform states.

Conclusion

The shift toward state educational financing in the past few decades has led to significant decreases in educational spending. Each state's finance system is unique, and its response to litigation related to education is different. However, research suggests that, on balance, court-ordered reforms have had their intended effects, reducing disparities in spending across districts through directing new state funds to low-spending and low-income districts. Additionally, preliminary evidence suggests that new resources have found their way to classrooms and that student achievement has not been harmed by court reforms.



The Economics of Grade Retention

Eric R. Eide, Brigham Young University

Grade retention is a widely contested issue in American education reform. Designed to improve students' academic skills and to allow them to mature, retention has been praised as an antidote to the social promotion of students who lack mastery of essential skills. However, empirical results on retention's benefits are inconclusive, failing to justify widespread implementation. The extent to which the policy improves student outcomes needs further economic assessment. This study reviews the issue's background, discusses the costs and benefits of retention, examines previous methods of estimating its effectiveness, and suggests a promising econometric estimation method.

Background

Repeating a grade is consistently associated with being male, minority, and poor. Students who are retained are often emotionally immature and younger than their classmates. Retention is common: An estimated 13% of youth aged 16 to 24 have repeated a grade, and 1% of K–12 students are retained annually. While some evidence indicates that retention improves academic achievement, self-esteem, and other outcomes, more evidence suggests that repeating a grade can cause low self-esteem, poor academic performance, and dropping out. As this study suggests, it is possible that the variety of data and analytical tools accounts for the preponderance of negative evidence, and perhaps more sophisticated analysis would reveal benefits more clearly.

Economic Effects of Retention

COSTS

Researchers typically focus on the benefits of retention. But they overlook the costs of retention, with the exception of emotional costs and

the labor-market penalties of dropping out because of retention. For thorough analysis of the effects of retention, three types of cost should be considered.

First, *the direct cost to the taxpayer* is large, estimated at \$2.6 billion annually at a conservative 1% annual retention rate. Teacher training and supplemental instruction for retained students are also costly.

Second, retained students potentially lose a year's income because of their *delayed labor-market entry*. If retained students graduate from high school and earn as much as other graduates, they may lose over \$30,000 for the retention year. Furthermore, if retained students drop out, they may gain earnings by early market entry, but these additional earnings may be offset by the lower lifetime income associated with failure to complete high school. Of course, if the retention is early and related to immaturity, such substantial economic costs may be compensated for by further education and job achievement.

Third, *psychological costs*, such as low self-esteem, may be either causes or effects of retention. These need careful consideration, given the mixed evidence on whether retention allows time for emotional maturation or inflicts emotional harm by creating or reinforcing a negative academic self-image.

BENEFITS AND THE DECISION TO RETAIN

In order for the measure to work, the benefits of grade retention must outweigh its costs, but little evidence exists that they do. Few have found that retention improves achievement, maturity, or post-school earnings; any advantages found tend to diminish with time. Even if better measures developed through future research show more benefits, the quality of the decision to retain on the basis of costs and benefits needs to be con-

sidered; an economic model may work best to that end.

The decision to retain is usually made by teachers and principals with parents, who may have the final say. The school's objective is always the same: to improve student performance. However, an economic model accounts for the parents' objective: that benefits in performance and expected earnings trump costs of deferred market entry and psychological damage. Since parents know good performance in a grade predicts success in later grades, they know their children's economic future depends most on performance in earlier grades; thus they have a greater incentive to retain their children earlier. Analysis of retention effects must account not only for the negative correlation between retaking a grade and later earnings but also for such endogenous—situationally specific—choices that could make that correlation positive.

But although the decision may be privately beneficial, it may be publicly negative, since most children are publicly educated, and parental decisions ignore the public costs of educating retained children another year. Since parents do not bear the full cost of retention, it is likely to have an inefficient outcome. And insofar as teachers decide, the outcome is inefficient in a different way. Since teachers receive no benefits from retention—indeed it costs them extra effort—they have an incentive to promote when they should retain for optimal student performance and future success.

Even when retention is based on a strict academic performance standard theoretically derivable from an objective cost-benefit analysis, the results are inefficient, because the standard originates in the school district, which tends to overpromote, cutting immediate reeducation costs at the expense of long-term outcomes.

An economical decision-making policy would have to change the incentive system so that decision makers faced the full costs and benefits of retention.

Estimating the Effects of Retention

Various types of data and analysis have been used to measure the effects of retention. An econometric approach may permit more informed decision making, since it controls for factors not easily observed yet having significant effects.

DATA

Most research on grade retention has used data from school districts. These data yield details about retention policy that are helpful in judging effects, such as the extent of supplemental programs. However, small sample sizes usually limit the generalizability of the results. Researchers have also used national surveys that are limited by their short-term nature or by their lack of data about reasons for retention, supplemental support, or long-term attainment.

ANALYSIS

The diverse analytical tools used partly account for the divergent results of retention research. Several common techniques have inherent advantages and drawbacks. First, qualitative approaches describe the results of student surveys, offering subjective insights not captured in quantitative studies. However, these approaches are clearly limited as grounds for larger inferences.

Second, basic quantitative research reports correlation coefficients between retention and possibly related factors, such as dropping out. However, the absence of controls for other factors makes it impossible to infer causation from correlation. More sophisticated quantitative approaches compare retained and promoted samples having similar characteristics. But those features can account for only some of the influences on retention, and it is likely that some major

determinants of the retained group's difference from those promoted remain unanalyzed.

Other comparative approaches are also problematic. While comparisons of retained and first-time students in the same grade can yield results on the value of retention, they cannot determine whether retention itself or another factor like maturation explains the added value. And while comparisons over time of retained students to promoted ones of the same ages eliminate the maturation factor, the exposure of promoted students to more material biases performance matching in their favor. Comparison results can differ on the basis of the comparison used. In one study, same-grade comparisons showed retention positively affecting performance, while same-age comparisons did not. Finally, multiple-regression analysis of both retained and promoted students has been used to control for a large number of observable variables affecting retention. When the sample size is large, this allows for a detailed analysis of retention's possible causes and effects.

None of these approaches, however, accounts for unobservable, usually endogenous, factors that may correlate with the likelihood of retention and with student outcomes. Immaturity, for example, is not typically observed by researchers, so a multiple-regression analysis will measure both retention effects and immaturity, leaving the unobserved factor as a possible cause and biasing the estimation of outcomes. To overcome this difficulty, econometric analysis affords a control technique using instrumental variables. In this method, the researcher identifies a variable that is highly correlated with retention but that is not correlated with unobservable factors influencing an outcome of interest. This instrumental variable can therefore only affect the outcome through its correlation with retention.

The present author has used the instrumental-variable approach to es-

timate effects of retention on adult earnings by establishing a student's relative age in her class cohort as the instrumental variable. This variable is correlated with retention, because younger students tend to be held back because of immaturity, but arguably the variable has no long-term association with outcomes such as later earnings or with unobservables. Standard multiple-regression analysis showed that retention strongly correlated with both dropping out and lower earnings, but when the instrumental variable controlled for unobservables, in most cases retention had *positive* correlations with outcomes, though these were not statistically significant ones.

This is weak evidence for retention policy, but it suggests that evidence against retention may be flawed. The lack of strong evidence for benefits of retention does not mean that no such benefits exist. Studies using instrumental variables suggest that better estimation may show more advantages, as may further consideration of long-term economic outcomes and of differences in advantages between borderline and weaker sets of retained students.

Conclusions

Although it is not yet clear whether grade retention is beneficial, it seems clear that social promotion, implemented broadly, is inefficient. It confronts neither the school nor the student with short-term costs, but it could lead to the long-term costs of lower student achievement and lower labor-market returns that are like to result from poor academic mastery. A better policy might be to promote students on the basis of economic incentives, so that decision makers face full costs and benefits. Short of comprehensively implemented private education, any market-oriented policy that allows for cost-benefit comparison should lead to more efficient retention decisions and better outcomes for both students and society.



Teacher Quality: Its Enhancement and Potential for Improving Pupil Achievement

Susanna Loeb, Stanford University

Because of teacher shortages fueled by increased enrollment and ample opportunities for potential teachers outside education, teacher retention and quality are central to school policy debates today. Incentive policies have been adopted to attract young people into teaching. While intended to boost the quantity of teachers, the policies have also attempted to distribute the quality of teachers equitably by providing incentives for better teachers to choose disadvantaged schools. This article considers the importance of teacher quality for educational policy by examining the effects of teachers on student performance, the distribution of teacher quality and its likely causes, and the role of policy decisions, especially about salaries, in improving quality.

Effects of Teacher Quality on Students

Research has clearly shown that certain measures of teacher quality affect student achievement gains measured by test scores. For instance, one study found a large difference, approximately 50 percentile points, in achievement between students attending classes taught by high-quality teachers for three consecutive years and those taught by low-quality instructors. The quality of teaching had residual effects in later years. The most common method of assessing teacher quality in such studies is the value-added method based on a teacher's history of improving test scores.

This method defines an important component of student success, but it involves potentially troubling features. First, teaching assignments may distort the quality measures, since assignment to accelerated or challenged learners may affect test scores despite a teacher's skills, resulting in over- or underestimation of

quality. Second, value-added quality is often measured in environments that give teachers of some students more incentive to focus their instruction on the test. These teachers may produce greater test-score increases in their students, even though they are not of "higher quality" than other teachers. Value-added measures therefore may not represent relative teacher quality accurately and may best be used in combination with other teacher measures.

Less agreement exists on the effects of specific teacher attributes than on overall teacher effects. Some studies examining teachers' years of education and experience, degree level, and certification, for example, did not find that these factors consistently affected pupils' performance. Other studies, however, did find that measurable teacher characteristics, test scores, and college quality, among other things, did affect students' test-score gains. Though some current research comparing value-added quality to teacher attributes found little correlation between them, these results were preliminary. Easily measurable teacher attributes like test achievement may be associated with teachers' influence on students and thus be important for policy decisions.

The Distribution of Teacher Quality

Recent studies have documented substantial differences in teacher qualifications across schools, districts, and regions. In California, for instance, in 10% of elementary schools, over 30% of teachers lacked full certification, while in many schools all teachers were fully certified. Similar disparities appeared in New York: In 5% of schools, fewer than 3% of teachers had a master's degree plus 30 credits, while in another 5%, over 60% of teachers had reached that educational level. Fur-

ther, differences in teacher characteristics correlate with each other; schools with less qualified teachers in one attribute tend to have teachers less qualified in other ways. Schools are not trading one type of qualification for another but simply hiring less or more qualified teachers generally.

Teacher characteristics are related to school characteristics. Schools in cities, those with high proportions of minority students, and those with many children in poverty tend to have teachers less qualified as measured by certification, degree level, college quality, and test scores. For instance, one third of New York City teachers taking the state's Liberal Arts and Sciences certification test failed it at least once, while only 4.7% outside the city did. Schools with 10% more Black students had teachers who averaged 5.7 points (almost 1 standard deviation) lower on the NTE knowledge exam than teachers at schools with more White students. Within districts, teacher differences similarly correlated with students' minority and poverty proportions.

Causes of Distribution of Teacher Quality

Differences in average teacher qualifications across schools may be due to hiring on the basis of characteristics that we do not have measures of, such as the ability to work with specific groups of children. But it seems unlikely that districts choose to hire teachers with lower qualifications on all of the multiple measures available. Thus, the variation in average qualifications seems unlikely to be the result solely of district preferences for different teacher characteristics.

Another possible cause of distribution by qualifications is hiring efficiency; some districts may recruit less aggressively and offer fewer incentives. This may account for differ-

ences across neighboring districts but not for sorting within districts of worse teachers to urban, minority, and poor schools. Some sorting may be due to school politics, as parental complaints about poor teachers can send worse teachers to worse schools. Nevertheless, it seems unlikely that all the sorting observed is due to transfers.

A more likely explanation for teacher quality differences is teacher preference. Research has shown that teachers move to schools with high-achieving, high-socioeconomic-status students when possible. Schools with low-achieving, poor students have difficulty attracting teachers. In some cases, that may not be because teachers are reluctant to teach students at less advantaged schools but because the facilities, supplies, and collegial interactions there are unappealing. Such effects of poor working conditions make it hard to assess the strength of preference as a cause of teacher quality distribution, but in any case it is likely a strong factor in that sorting.

Salary Policy and Teacher Quality Distribution

Though intradistrict salary differences exist too rarely to play a major role in quality sorting, interdistrict starting-salary differences between high- and low-achieving or high- and low-poverty districts are substantial, with the districts with stronger, richer students paying thousands of dollars more, enough to motivate preferential choice. Though starting salaries vary substantially among metropolitan regions of the country, all regions show similar quality differences, so it is likely that cross-regional variations reflect competition from nonteaching opportunities more than differences in cross-regional quality that are related to sorting.

Much research has shown that teachers do respond to salary differences. Individuals are more likely to choose to teach where starting salaries are high in relation to other occu-

pations' wages, and established teachers are more likely to leave teaching when they work in low-salary districts. These findings contradict surveys of teacher choice indicating that motives like the social value of education are more important than salary. However, such findings overlook the common disparity between ideals and action, which makes it possible that salary is a more important factor than teachers admit.

Although high salaries can attract college graduates to teaching, they do not necessarily increase teaching quality. Higher wages will increase the applicant pool, but no direct evidence suggests that better wages draw better teachers disproportionately. The association of higher wages with incentives to work at poorer schools and with competition from alternate occupations also suggests that high pay and quality are not necessarily connected. However, a recent examination of national data since 1960 indicated that states and districts with increased wages saw the most gains in student achievement. This is a sign that high wages can attract teachers and that they may also indirectly improve student performance.

Conclusion

The literature provides evidence that schools differ starkly in the average characteristics of their teachers. It also suggests that teachers respond strongly to both pecuniary and non-pecuniary elements of their jobs. Across districts, in general, there are substantial wage differences. Some wage variation appears to contribute to district differences in alternative wages, but other variation appears to contribute to district differences in average teacher quality. Salary schedules differ little within districts, yet there is large variation in teacher characteristics among schools within the same district. Current salaries can be neither driving nor alleviating these differences. Variation in teacher quality appears to be driven primarily by

teachers' preferences for high-achieving, high-socioeconomic-status students or by working conditions in these students' schools. Targeted salary increases and targeted improvements in working conditions in difficult-to-staff schools are needed to draw high-quality teachers to low-performing schools and to alleviate disparities in the quality of the teaching workforce.

Teacher preferences have created disparities in schools' ability to attract and retain high-quality teachers. Teachers, like most workers, value high salaries. However, they also care about the types of students they teach and the environments in which they work. While schools can all pay the same amount for a textbook or a computer, the same teacher will cost different schools different amounts. If schools with unattractive working conditions wish to hire and retain teachers of the same quality as teachers at other schools, they will need to improve these conditions or pay higher salaries.



Look for the next
issue of the CEIC

Review:

*School-to-Work:
Accomplishments,
Problems, and
Prospects*

Measuring School Efficiency: Lessons from Economics, Implications for Practice

Amy Ellen Schwartz and Leanna Stiefel, New York University

Estimating efficiency and productivity in education involves confronting and addressing a host of difficulties in measuring input and output, capturing environmental influences, compensating for data scarcity, and determining causality. Nevertheless, recent improvements in data quality and availability and accompanying advances in statistical methods offer the promise of improved measures of school efficiency and the prospect of identifying the determinants of efficiency across schools and school districts and over time. This report discusses approaches to measuring K–12 efficiency and the relative merits of each. It also explains the complexities of inputs, outputs, and units of analysis before indicating lessons learned.

Measuring Efficiency

Economic efficiency at the school level means that schools produce the most education possible with given resources, use the least expensive combination of resources for any given level of education, and produce the level of education that consumers prefer. Most research has focused on estimating production relationships in education, formulating a production function that measures the maximum output that a given quantity of inputs can produce. This relationship can be statistically estimated and the results used to identify schools' relative efficiency.

Regression estimation of production functions can yield school efficiency measures in two ways. First, school fixed effects can be estimated. These capture the extent to which a school's output systematically exceeds (or falls short of) the output of the average school, given the variations in inputs, students, and their peers' characteristics. Thus, this fixed effect can be viewed as a measure of the relative effi-

ciency of the school in comparison to other schools in the sample. Second, efficiency can be gauged using the estimates of the marginal impact of resources on outputs from the regression along with input prices. These provide insight into whether additional inputs lead efficiently to additional outputs and also into whether greater outputs can be produced by reallocating existing resources.

While most school efficiency work is based upon such regression estimation, three noteworthy alternatives are available and offer a variety of advantages, although these are not without their own drawbacks. Most attractive is that these alternatives incorporate more than one output measure, thus addressing an important limitation of much efficiency research.

First, *cost-function-based efficiency estimates* capture the relative cost of each school's production of education, controlling for different outputs and prices. Second, *property-value-based measures* gauge efficiency on the basis of the extent to which school quality and school taxes are capitalized into property values. This method can yield estimates of the relative importance of, say, test scores and graduation rates for taxpayers. Finally, *data envelopment analysis* (DEA) encompasses programming models that estimate efficiency by constructing a nonstochastic production frontier of greatest efficiency from observed inputs and outputs. With DEA, several outputs can be considered simultaneously, and in addition to distinguishing the relative efficiency of the school, they can provide insight into reductions in inputs that could be accomplished without a reduction in outputs.

Measuring Efficiency in Practice

SPECIFYING INPUTS

Specifying appropriate measures of inputs is critical to efficiency

measurement—bias can be introduced either by omitting inputs to production or by including irrelevant variables. Four input categories are relevant to productivity. Some, schools can change; others are relatively immutable.

In the first category are the *purchased physical inputs*, including teachers, supplies, and buildings. While these are the inputs administrators might be most likely to consider in trying to increase output or efficiency, relatively little school-level data on physical inputs are available—typically teacher counts and spending information. Unfortunately, spending reflects both prices and quantities, rather than just quantities, which would be preferable.

Second are *donated physical inputs*: expert and parent volunteers and donated materials like computers and supplies that can improve productivity. While typically welcomed by the school, for a variety of reasons these inputs may involve opportunity costs, and their availability can vary widely from school to school, depending on the resources and characteristics of the school and its community.

The third, varied group is *individual student-specific inputs*, including students, families, and their productivity-relevant qualities. Outcomes depend on factors like student motivation and parental education as well as school factors. Since schools can affect these inputs, for instance English proficiency, by selection processes, separating production efficiency from selection is challenging. It requires detailed, longitudinal data, which are often scarce. Nevertheless, progress in gathering such data is helping researchers disentangle these.

Fourth, *peer and community environment* can determine school efficiency, influencing student characteristics and other inputs. Student performance is likely to be influenced

by community, cultural, and economic factors, which must be carefully separated from the influence of the students on each other.

SPECIFYING OUTPUTS

Educational outputs are harder to specify, partly because so much research has relied on models using one output. Despite the need for the multiple-output models mentioned earlier, one-output models are simple, and since alternative outputs are typically highly correlated, some argue that one is sufficient. Test-score achievement is the most common output measure used, largely because scores are available, providing consistent, quantitative measures suitable for this sort of analysis. Notice that different tests may yield different efficiency measures for a given set of schools, depending, for example, upon the subject of the test (e.g., reading versus math), or the scoring (e.g., norm-based versus criterion-based tests).

In addition, if the measures are used in decision making, for example, to reward or recognize performance, different measures may provoke different responses. For instance, reliance upon a measure of the percentage passing a test will provide an incentive to focus efforts on helping weaker students pass, not on improving the performance of the stronger students. The choice between output measures cannot, then, be determined statistically but must reflect the purpose to which the efficiency measures will be put.

Other output measures include graduation and promotion rates, but inconsistent standards governing their definitions render them problematic. A popular and useful formulation of output relies upon value-added output measures, which ideally are measured as the change in a given group's performance between periods. Unfortunately, these data are typically unavailable, so changes across school grades are often used. This is particularly problematic in

urban and other areas in which school mobility is high.

UNITS OF ANALYSIS

While schools are one logical unit of analysis, since they have sanctioned responsibility for efficiency, moving below the school level to students or classes is attractive for a variety of reasons. Perhaps most important is that it offers the opportunity to understand differences in education across grades and groups of students. Grade-level analysis may prove more rewarding by avoiding inappropriate comparisons of differing school structures (for instance, K-3 and K-6 elementary units), by identifying efficiency differences across grades, and by discovering the most efficient grade organizations.

Unfortunately, output data will be easier to obtain than input data, since individual student test data are typically recorded, while input flows are rarely tracked within schools. Though trends towards teacher accountability call for class-level analysis, this kind of investigation is currently too costly. Group analysis by race or gender, for which SAT and other data are available, may be useful, but the most challenging and fruitful work may be at the student level.

A rich specification for efficiency estimation can be based on student-level data, which allow detailed accounting for differences in socioeconomic, schooling, and demographic characteristics among students. For instance, sharper analysis of specific effects of peer groups on achievement is possible through the use of a student-level model, and the effects of student heterogeneity can be better measured.

Commentary and Conclusions

Several unresolved issues related to school efficiency remain. First, although richer, more detailed data sets are available for analysis, the quality of the data is unclear. For example, the common measure for

poverty, federal school-lunch eligibility, may prove to be an insufficient measure of student poverty, particularly for high-school students. Capital and building statistics, moreover, are incomplete, and yearly test scores may not measure what we mean by performance. Conceptually, the assumption that all schools attempt to maximize productivity and outputs needs further investigation. Output equations may represent only relative, not optimal, productivity, if the sampled schools are off the production frontier.

Further, the extent to which existing analyses are limited by omitted-variable bias remains unresolved. Endogenous relations, such as inverse links between government aid and performance or between class size and performance, may seriously confound the estimation of production functions. Identifying instrumental variables to avoid endogeneity has been difficult. Further, defining an appropriate set of schools for multiyear analyses will require difficult decisions to be made. Omitting schools that exit or enter during the study period may introduce bias, as it may fail to account for school reorganization or redefinitions. Finally, additional research investigating the measurement of school efficiency in theory and in practice, comparing the alternative methods, and assessing the quality and contribution of new data, is critical.

This is an exciting time for research on education efficiency, as better data gathering keeps pace with theoretical advances and increases in computing capacity. The lessons of existing research, however, are important. First, efficiency measures can provide useful and needed insight to guide decisions about school improvement, and second, the employment of multiple measures is preferable to relying on a single, "definitive" measure to capture school performance.



Examining School-Level Expenditures and School Performance: The Case of New York City

Ross Rubenstein, Georgia State University; and Patrice Iatarola, Institute for Education and Social Policy, New York University

How school resources affect student outcomes has been the subject of long-standing debate. A common problem in determining how resources help students has been the difficulty identifying how resources reach students, since analysis has focused on school-district data. With improved data collection technology, though, researchers have begun to analyze resource allocation at the school level. As districts decentralize and schools increasingly control their own budgets, meaningful allocation differences among schools should emerge. This article examines school-level spending patterns in New York City public elementary schools. First, earlier research on resource allocation is reviewed. Then data and methods are explained, high- and low-performing schools' allocations are analyzed, and future research directions are suggested.

Earlier Research

Since schools produce educational outcomes and actually expend resources, basing accountability and policy in schools rather than districts is a logical progression, and school-level research seems most useful to inform future policy. However, data at the school level are often incomplete or inaccurate. Moreover, access to school data is difficult in states where information is aggregated at the district level. Further, states vary widely in collecting and reporting methods. Conceptual difficulties also arise through unclear staff-time accounting, varying concepts of school units, and inconsistent definitions of functions and objects such as teaching positions.

Despite such problems, some states and cities have made efforts to improve access to school-level financial data, though more state efforts are needed. Findings regarding school spending patterns have been

surprisingly consistent. Most studies find that schools spend approximately 60% of their budgets on instruction with little variation in relation to levels of per-pupil expenditure or school size. Comparisons of instructional and administrative expenditures to investigate possible diversion of classroom resources have shown that high-performing urban schools often allocate less discretionary spending for administration than do weaker schools. However, combined administrative expenditures in a New York State study were low enough, about 11%, to conclude that such allocations have little effect on student outcomes.

Increasingly, investigations have examined how schools can reallocate existing revenues to support reforms. Additional resources have been largely dedicated to hiring additional staff, independent of the school's budgetary discretion; more fundamental reforms like class-size reductions are usually initiated with federal categorical funding. Another concern has been allocational inequities. These are less likely to appear in intradistrict comparisons, given district tax-base uniformity, but such comparisons have shown that teachers with less training and experience are more likely to be found in urban schools serving primarily students from low-income families. Districts often allocate positions rather than salaries to schools, which may exacerbate this problem.

Data and Methods

The New York City public school system is the largest in the country, with over one million children and over one thousand schools. With full budgeting and instructional authority, 32 subdistricts govern the system's presecondary schools. The Central Board of Education allocates resources to the subdistricts through formulae for teaching positions and

other needs. However, the system is changing dramatically through its new Performance Driven Budgeting initiative, which links budgeting decisions at the school level to instructional decisions with the goal of improving performance. Begun in six strongly managed subdistricts, the initiative in its first two years has shown success, including earlier allocation decisions and computerized budgeting implementation.

The authors used the city's data on expenditure, teacher and student characteristics, and student reading and math achievement for public elementary schools. The rich expenditure data were classified by type of expenditure, educational setting, and funding source. City, state, and federal funding sources were specified. Data were collected on over 600 schools for 1997 and 1998. Students were typical of large urban districts. They scored around the national median on reading exams, were 75% poor (according to federal school-lunch eligibility), and 16% had limited English proficiency. Average per-pupil expenditures rose from 1997 to 1998.

These data were used to select subgroups of high- and low-performing elementary schools and to compare their allocations of resources. Schools were judged on the basis of increases in the proportion of their students scoring above the national median on the exams. The authors selected schools through a value-added model, using the change in the percentage of students who scored at or above the 50th percentile on the reading test from the previous year's third grade to the current year's fourth grade, and controlling for socioeconomic variables.

The difference between schools' expected performance (as predicted by socioeconomic characteristics and previous test scores) and their actual performance was used to rank them

into quartiles. Those schools that appeared in the highest and lowest quartile for both 1997 and 1998 were then selected for the purposes of analysis as higher and lower performing schools, of which there were 34 and 37, respectively.

Results

The value-added regressions yielded significant findings. First, the model accounted for 75% of the variation in reading performance. Second, immigrant status was, surprisingly, positively associated with performance, while English proficiency had the expected negative association. Perhaps immigrant families cluster near better schools. Third, poverty was, also surprisingly, not significantly related to performance, possibly because of the close relationship between third-grade test scores and poverty.

Comparison of the two school groups showed them to be similar in socioeconomic characteristics, such as poverty and the percentage of students with limited English proficiency. However, analysis revealed large differences in student reading gains between third and fourth grades. In low-performing schools, the percentage of students scoring above the 50th percentile decreased between third grade and fourth grade in both years, while the high-performing schools showed large average gains between the two grades. And while in both years the weaker schools moved from slightly above the city reading average in third grade to just below the median in fourth grade, the stronger schools moved from well below to well above the median from third to fourth grade.

Differences also emerged in financing. While both groups spent more per pupil than the city average, lower performing schools controlled less of their spending, and they received less of their funding from city and state aid. Key differences appeared in teacher characteristics. While both groups spent similar pro-

portions of their funding on teacher salaries, the stronger schools examined exceeded the city average in teacher salary, experience, education, stability, and qualifications, while the weaker schools examined proved below average in these characteristics. These results suggest that teacher distribution may be related to student performance.

The pattern does not indicate causality, however. Since city union rules favor senior teachers when positions open, senior teachers may go to schools with higher performing students. This explanation is unsatisfactory, though, because both school groups were similar in student socioeconomic characteristics, and higher performing schools showed lower third-grade performance. Still, exogenous characteristics of students and schools may run counter to practices of resource allocation in attracting teachers. Without further research, no rigorous conclusions can be drawn about the effects of teacher qualities on student performance, but these results suggest that better teachers, not just more funds for teachers, are important to student success.

Further significant patterns appeared in other spending functions. Most of the budget was spent on direct services in all schools. Over 80% of the resources spent on direct services was spent on classroom instruction. Because most of this funding was directed towards personnel, there was little room for flexible spending. Instructional support and leadership allocations, the next largest budget share, varied little from school to school. However, while both school groups dedicated less than 1% of school-level budgets to professional development, high-performing schools allocated twice as much per pupil. Stronger schools also dedicated more per pupil for paraprofessionals and for instructional support, while weaker schools spent more per pupil on teachers.

More differences in proportional spending appeared in the spending

that was controlled at the subdistrict level. The subdistricts of weaker schools spent more for direct services and instructional support and less for professional development and leadership than did the subdistricts of stronger schools.

In a comparison of the two groups in base funding from the city and state operating aid, minimal differences emerged. But in other funding from federal programs like Title I and from other state and private sources, significant differences became evident. The lower performing schools allocated about \$50,000 per school less of such funding to paraprofessionals, instructional support, and professional development, while allocating more for teachers. Half this difference stemmed from higher spending for instructional support in the higher performing schools.

Conclusions and Future Directions

School-level financial reporting in New York City provides rich detail on spending but cannot definitively relate financial information to educational outcomes. Moreover, the lack of variation in spending across schools makes it hard to identify expenditure patterns that definitively improve achievement. Further, the lack of expenditure differences between stronger and weaker schools is surprising. Another notable finding from this research, furthermore, is the greater variation in spending patterns found for resources that schools themselves control, such as those from Title I.

Future research could explore disparities between high- and low-achieving schools in different test subjects, grade levels, and years examined, ideally using student-level data for more precise analysis of achievement changes. More microanalysis of expenditure patterns is needed to detect variations in expenditure that school or district averages obscure.



The Relationship Between Student Performance and School Expenditures: A Review of the Literature and New Evidence Using Better Data

Corrine Taylor, Wellesley College

The questions whether and how much additional school spending improves student outcomes remain essential to policymakers allocating limited resources to schools and other public programs. These questions have been investigated for over 30 years, and the results of econometric studies of “the educational production function”—of effects of school inputs on student outcomes—vary widely. Positive, inverse, and no effects have been found. Major syntheses of studies have also reached opposing conclusions.

This article reviews the educational production-function literature to explain the varying results. It then describes how new, combined data sources allow researchers to examine three overlooked empirical issues related to conflicting findings. Finally it reviews the author’s analyses of those issues and relates new findings to earlier research.

Literature Review

DIFFERING FINDINGS

The many production-function studies have differed, often because of data variation, in the following five ways. First, studies varied in the *choice and measurement of outcomes and input variables*. Educational or labor-market outcomes were examined as outcomes, and school expenditures, teacher salaries, or class size were input variables of interest. While researchers preferred to examine as many family and peer control variables as possible, no consensus arose on the variables’ importance or measurement. Second, the *unit of analysis* differed, despite consensus that analysis of student-level outcomes and spending would be ideal. In reality, school resources were usually reported at higher levels, from classroom to state, constraining studies to these variously compiled inputs.

Little consensus developed on the third difference, *geographic scope*. Nationwide studies offered superior input variation, but inability to control for different state policies resulted in omitted-variable bias. Favoring nationwide scope, some studies criticized statewide studies for endogenous biases. A fourth difference arose in studies’ *functional form*. Earlier studies assessed achievement once without considering prior influences, but later studies controlled for earlier achievement, isolating the value added of the inputs. Finally, analyses employed various *estimation techniques*. Most used least squares regression, but recent studies incorporated advanced econometric techniques to address problems like selection bias. These differences have contributed to disparate results on the relationship between student outcomes and school inputs. Several syntheses combining results of primary studies have addressed whether any systematic relationship exists.

SYNTHESES

The earliest syntheses concluded that school resources and performance were not consistently linked. However, the synthesis method used—vote-counting—has been widely criticized. More sophisticated meta-analyses using the same pool of studies have recently shown positive and substantial links. Comparison of these methods can help evaluate their merits.

The vote-counting syntheses categorized the relationship between input of interest and outcome for each study as positive, negative, or unknown and categorized the relation as significant or not. These “votes” yielded a percentage distribution of estimated direction and significance for each input of interest. In one vote-counting synthesis, though most studies (61%) showed positive rela-

tions, only 27% were statistically significant, so it was concluded that there was no strong relation. Yet no rules determined what proportion of significant positives was necessary to support a systematic, positive relationship. Moreover, vote-counting overlooked sample and effect sizes, which affect statistical results. The method was not statistically sophisticated enough to support or reject hypotheses about associations between inputs and outcomes.

More recent meta-analyses have used two more sophisticated methods: combined significance testing and effect magnitude estimation. Combined significance testing calculated probability values from studies weighed by sample size to test significance for school inputs. All inputs tested showed positive coefficients, and the majority showed no negative coefficients, suggesting that the overall input–outcome relationship was positive.

Effect magnitude estimation measured effect size, indicating the strength of the relationship between a dependent variable—achievement test scores—and specific independent variables across studies after rescaling to align variables measured differently. Positive coefficients were found for each input in the full analyses and in study subsamples. The median effect size for per-pupil expenditure (PPE) showed a \$500 increase in expenditure associating with a large but plausible achievement increase of .15 standard deviation. In sum, more sophisticated syntheses tend to support positive school effects, and better data becoming available facilitate better primary analyses.

New Data and Evidence for the Relationship Between Expenditure and Achievement

Though policymakers are concerned about the relationship be-

tween PPE and achievement, studies of PPE effects have been hindered by aggregated expenditure analysis without family background data and value-added factors. However, combinations of relatively new data sources allow researchers to analyze three crucial production-function issues not explored earlier: (a) interdistrict variations in education costs; (b) the endogeneity of students' decisions to stay in high school; and (c) production-function differences in varying school settings.

To analyze these issues, three national databases from the National Center for Education Statistics can be combined. First is the National Education Longitudinal Survey of 1988 (NELS), which is a nationally representative sample of eighth graders who were tested and surveyed in 1988, 1990, and 1992. A second data source can be added to NELS: the Common Core of Data (CCD), which includes 1990 to 1992 district financial data and 1990 census data, including special-needs and community information, tabulated for districts. A third source that can be combined with these is the Teacher-Cost Index (TCI), which measures variations across districts in nondiscretionary teacher resource costs, using teacher survey data from 1990 and 1991.

Merging these data creates a panel of 14,000 public-school pupils with detailed student, school, and expenditure information. The author has used this panel in regression analyses to examine the three overlooked issues described above. Student math test scores from the 1992 NELS were the outcome of interest and PPE the explanatory variable of interest. Control variables for eighth-grade achievement measured gains, and other variables measured the influence of student, peer, school, and community features.

ANALYSIS 1: RESOURCE COSTS AND STUDENT NEEDS

Since resource costs (notably teacher compensation) and the pro-

portion of costly special-needs students can vary substantially, overlooking these factors can lead to underestimation of the link between expenditure and achievement. Thus controlling for such fiscal variables should increase the effect of PPE on achievement. The author tested this hypothesis, using cost-adjusted PPE and controlling for student proportions needing special education, English-proficiency support, and compensatory education. Positive relations between PPE and achievement were found. Coefficients were consistently small, and relational increases were too small for these fiscal factors to account for the weak expenditure-achievement relationship in the literature.

ANALYSIS 2: ENDOGENEITY OF EDUCATIONAL ATTAINMENT

Few input-outcome studies have measured the endogeneity of secondary students' educational attainment decisions. Dropouts are simply excluded from most studies. But when school-input effects on the decision to stay are overlooked, the research suffers from selection bias. If we assume that higher achieving students are more likely to stay in school and that PPE correlates with the proportion of students staying in school, then the effect of expenditures on achievement should be statistically stronger for all students than for those who stayed considered separately. To test this hypothesis, the author used the NELS test scores of dropouts in a two-stage selection-correction procedure to estimate direct and indirect (through attainment decisions) effects of PPE on achievement. Findings here showed that correcting for the endogeneity of attainment decisions only slightly increased the effects of PPE on achievement.

ANALYSIS 3: DIFFERENCES ACROSS SCHOOL SETTINGS

Researchers have documented striking differences across school settings in resources and student,

teacher, and community characteristics. The panel data revealed that inner-city and poor rural students were educationally disadvantaged in family and community resources. Additionally, school spending for these students averaged about 5% less than spending for students from suburban, White, and advantaged (SWA) schools. Test and retention outcomes were also lower for disadvantaged students.

A deeper question is whether educational production functions themselves differ across inner-city, poor rural, and SWA schools. Separate production-function studies were conducted for the three distinct school settings, and the magnitudes of the coefficients on PPE were compared. Findings revealed that PPE effects on student achievement varied significantly across the settings. Inner-city effects were consistently small, while poor rural and SWA students enjoyed larger, positive, and significant effects close to the .15 standard-deviation level and much larger than the effects in the resource and attainment analyses.

Conclusion

This new contribution to the educational production-function literature indicates a positive but relatively small effect of school expenditures on student achievement. Improvements in addressing differences in resource costs, special-needs student populations, and student retention produce small changes in the measured effect of PPE on student achievement. The benefits of conducting separate analyses for different school settings appear to be greater. The main policy implication of this research is that simply equalizing PPE for similar student populations is insufficient to produce similar student outcomes. Further effort must determine which purchased resources are most effective for different students and settings.



Exploring the Nexus Between Educational Inputs, Distribution of Performance, and Incentives

Samid Hussain, Cornell University and Cornerstone Research

In response to the current movement to improve performance standards in public schools, policymakers have been examining more carefully the productivity of educational inputs, whether of funding, teacher quality, or incentives. Research on productivity (known as education production-function research in the literature), most of it identifying the effects of inputs on student achievement tests, has yielded mixed results, largely because of disagreements about the best ways to measure productivity.

Conflicting evidence on the degree to which inputs improve performance has fueled debate on allocation of resources to schools. Research has called for more rigorous examination of inputs like class size, teacher test scores, and teacher education. A key assumption of past research, that only average test scores of a school's students matter, has overlooked potential effects of various inputs on the distribution of test scores.

Examining these effects might help policymakers aim resources and incentives at benefiting those low- or high-achieving students whom inputs may influence more. Some evidence has shown that increased spending benefits low-performing students more than it benefits other student groups, so focusing attention only on average scores might be misguided.

Building on earlier research, this study advances past analyses of education production functions in three ways. First, the study explicitly links teacher incentives offered by schools to productivity of inputs, specifically by examining schools that do and do not recognize teachers' accomplishments. Second, it examines how resources and incentives impact the entire distribution of student performance in schools. Third, it investigates whether improvement in mean performance is associated with im-

provements in performance of low performers or whether improvement in mean performance takes place at their expense. Discussed are the economic optimization model underlying the study and the data, econometric method, results, and implications for policy reform.

Propositions from an Optimization Model

The study begins with a theoretical model linking educational inputs to the distribution of performance under alternative incentive types. The model focuses on the central administrative goal of learning by all students and on issues of motivation at the decentralized level. The theoretical model distinguishes between the impact of implicit salary incentives based on experience and education and that of explicit salary incentives based on student performance.

This theoretical model leads to three propositions. First, effort at decentralized levels is consistently high only where incentives are explicit. This result may reconcile mixed evidence in the existing literature regarding the relationship between performance and educational inputs when key student characteristics are controlled for. Second, an increase in dispersion in performance associated with an increase in the mean performance will be less dramatic in an explicit incentive regime than in a regime offering few or no incentives. Third, a salary-based incentive scheme may be a very costly approach to motivating teachers and school administrators to put forth high effort.

Data

The study uses the National Educational Longitudinal Survey of 1988 and its two-year follow-up study of a range of data from nationally representative samples of 8th and 10th graders. The sample analyzed here was

limited to 10th-grade public-school students who took a mathematics achievement test (IRT) in the 8th and 10th grades. In order to analyze the effects of inputs on performance, 3,500 students were sampled, while for examining mean–variance trade-offs at the school level, 7,750 students in 811 schools were studied.

Also considered was information on whether or not schools offered recognition of teacher accomplishment like load relief. Such recognition was used as a proxy for teacher incentives.

Method

The econometric estimation of the relationship of inputs and performance proceeded in two steps. First, the standard value-added specification was adapted for school incentives in order to investigate the determinants of student performance together with school incentives. This step permitted testing whether the effects of selected variables differed between schools with or without incentives. Second, the question of how the benefits from higher average performance were distributed among students was investigated. Here, the dispersion in performance at the school level was operationalized by the standard-deviation measure, while the arithmetic mean of the student performance was used to describe the average performance. All relationships were analyzed by using both least squares and simultaneous quantile regressions. Detailed analysis and tabulation of these regressions appear in the book, *Improving Educational Productivity*, resulting from the conference summarized in this issue.

Results

The first major finding from this study was that educational inputs and incentives differentially impacted the student performance distribution.

More specifically, school inputs mattered more for the low performers and were frequently more productive in schools that offered some incentives. With regards to the teacher characteristics, the majority of teacher variables were not statistically significant in schools that did not recognize accomplished teachers. For example, the amount of teacher experience was relatively more productive in schools with incentives in terms of both the mean effect and the effect at the lower part of the performance distribution.

The second interesting finding pertained to contrasts between those teachers who had standard certification (relative to provisional or no certification) and teachers who had bachelor's majors in mathematics. Specifically, if a teacher had standard certification, then he or she was relatively more productive (in terms of the mean effects on student performance and the effects at the 5th and 75th quantiles) in schools that offered no incentives. In contrast, if a teacher had a bachelor's major in mathematics, then he or she was relatively more productive (in terms of the mean effects and the effects at the 5th and 75th quantiles) in schools that did offer incentives.

The third important finding was that higher starting salaries, controlled for experience and education, were associated with higher student performance in the lower part of the performance distribution (i.e., 5th and 25th quantiles) only for schools that recognized accomplished teachers. This effect suggests that, to the extent that those with the lowest salary teach the lowest performers, increasing the salary floor increases performance at the bottom of the performance distribution. Hence, incentives work.

Another major finding was that not all of the educational inputs were consistently productive in schools recognizing accomplished teachers. This suggests that the existing incentives may not be sufficiently explicit to render all of the inputs productive.

Such insights can allow policymakers to design interventions that target resources more efficiently and equitably towards low achievers and high achievers, while being attentive to the impact of incentives.

A further set of findings from this study pertained to the relationship between dispersion in performance and the average student performance. The key finding was that the relationship between dispersion in performance and the average performance was sensitive to the functional form estimated. This complication underscores the need for a more rigorous theory of potential trade-offs between mean and variance before a precise relationship can be ascertained. Also, an increase in mean student performance in schools that recognize accomplished teachers was typically associated with a less dramatic increase in variance.

Moreover, an increase in variance due to an increase in mean performance was less dramatic at the lower end of the standard-deviation distribution than at the upper end of the distribution in schools that recognized accomplished teachers. This suggests that such incentives are instrumental in restraining any increase in dispersion associated with the mean performance, especially at the lower end of the standard-deviation distribution. Other factors that differentially affected the dispersion in performance were school inputs such as school size, the lowest teacher salary in the school, the concentration of teachers with a master's degree, and the percentage of full-time faculty in math.

Conclusions

In the context of the public debate about standards and accountability as well as in the context of the research debate about the effects of education inputs on student performance, it has become increasingly crucial for policymakers to understand how policy-relevant, resource-related inputs are associated with productive

student outcomes, so that they can allocate resources more effectively. This study's findings are relevant to that understanding.

The author's regressive analyses of national educational-survey data have shown that, at least in relation to mathematics performance outcomes, school inputs matter more for low performers and are frequently more productive in schools that offer teacher incentives. This suggests that policymakers should consider allocating more resources toward recognition of teacher achievement. Furthermore, incentives have been shown to be instrumental in restraining any increase in dispersion associated with an increase in mean performance, especially at the lower end of distribution. This finding also supports the use of resources for teacher incentives, since it suggests that recognizing teachers can have an equalizing effect, especially for weaker students.



In a future issue
of the *CEIC Review*,
find results from a
national invitational
conference on
class size reduction:
*Taking Small Classes
One Step Further*

Problems in the Estimation of School Effects: Insight from Improved Models

Jens Ludwig, Georgetown University

Analysts agree that educational reform could be more effective if changes could be found that would generate better outcomes for a given level of school spending. However, determining the causal relationships between interventions and effects, which is needed for successful policy, is difficult. Differences among students can affect outcomes as much as inputs, and these differences are hard to isolate. This article examines these causal relationships, describing the evaluation problem in outcome analysis, explaining nonexperimental approaches to the problem, and assessing how well the approaches work and what we have learned from them. Standard approaches can yield misleading results and lead to ineffective policy.

The Evaluation Problem

Evaluating educational interventions is complicated, because we only observe outcomes that students experience for the interventions to which they were actually exposed, but not potential outcomes from alternative interventions. Therefore, identifying the differences that policies make is difficult. Program evaluation analysis thus uses statistical methods to estimate outcomes with and without exposure to interventions. Randomized experiments, by providing control groups equivalent to treatment groups, allow comparison between outcomes with and without interventions.

Unfortunately, since support for experiments is weak, analysts must rely mostly on nonexperimental comparisons of students who receive an intervention with those who do not, hoping that statistical methods can account for all the factors beyond the intervention responsible for different outcomes. The challenge for analysts is accounting for the many in- and out-of-school factors that may affect selec-

tion for intervention and achievement but which cannot be readily measured.

Researchers usually assume that educational inputs produce outcomes in linear-equation fashion: A set of measurable input factors amounts to the outcome without regard to differences between students, and a change in any factor will have the same effect on an outcome. However, students who receive inputs of interest may differ systematically from those who do not because of unobserved factors, causing analysts to confuse the input with the unobserved factor as the cause of the change in outcome. Self-selection and school targeting into programs are factors that are hard to observe in large-scale analyses. Many analytical approaches have tried to overcome this evaluation problem.

Nonexperimental Approaches

Most educational studies through the 1970s used a cross-sectional approach. Relying on data sets measuring students' background, school inputs, and test scores at one point in time, researchers hoped to gather enough data to account for all causes of student differences. However, unmeasured reasons for choosing or being assigned a given input have severely limited this standard approach.

The 1980s and 1990s saw an improved approach, represented by the U. S. Department of Education's High School and Beyond (HS&B) survey, which sampled high-school sophomores and seniors over more than 10 years. Multiple observations allowed researchers to study achievement after controlling statistically for students' previous test scores, under the assumption that previous test scores would serve as a proxy for unmeasured background characteristics affecting achievement. However, this value-added approach

still left causal factors unaccounted for, such as test-score growth rate or factors intervening between test periods—a significant two years in the HS&B study.

Another approach to accounting for unobserved factors is using data from multiple children in the same family to control for unmeasured family variables, which are assumed to be constant and to affect all siblings equally. Therefore, differences between siblings receiving and not receiving an educational intervention are assumed to be due to the intervention and other measurable inputs. However, unmeasured differences between siblings may still bias estimates.

A further method for considering student differences that can bias simple statistical models is the propensity-score method. Here, student background data are used to estimate sample members' probability (the propensity score) of participating in the educational intervention of interest. Those receiving the intervention are matched only with those without it who match their propensity scores; unmatched data are discarded. If the background information is full enough, the treated and comparison groups will presumably be similar enough to factor out unobserved influences on outcomes.

A different approach is the instrumental variables (IV) method, which strives to identify a factor explaining why some students receive an intervention, such as school vouchers, and others do not. This factor is unrelated—exogenous—to student characteristics affecting outcomes. The exogenous factor can be related to the outcome of interest separately from background characteristics in a two-stage regression analysis. If the IVs are indeed exogenous to unmeasured characteristics and do explain reasons for intervention receipt, they can yield more accurate

estimates of outcome. Like randomized experiments, the IV approach can identify and isolate factors affecting intervention so that choice for intervention is uncorrelated with specific aspects of the students examined. When the IVs identified are valid, this approach can be a powerful tool for overcoming self-selection bias. At the least, the method focuses analysis on reasons for student differences needing consideration.

A related approach, the two-stage selection–correction model, calculates distribution of estimated inputs to arrive at a nonlinear proxy for the unmeasured variables leading to selection bias. These calculations can be made even when valid instrumental variables are not available, but in that case the statistical validity of the proxy is affected.

Assessment of Approaches

How well have these approaches worked, and what has their application shown about educational policy effects? Value-added analyses, though highly regarded, have judged the effects of interventions poorly. Empirical testing of their application to the rich National Education Longitudinal Survey (NELS) of 1988 indicates considerable bias. Comparison with a randomized class-size experiment shows that the value-added model, based on incremental gains, cannot explain the large one-year gains appearing in the experiment.

Nonexperimental methods have generally been unable to reproduce experimental estimates, as comparison of experimental and nonexperimental data in adult job training has shown. More recently, systematic testing has permitted identification of nonexperimental models that match experimental estimation more closely. A recent comparison of a randomized housing-mobility experiment relating relocation to welfare benefits showed a good match between experimental and value-added results. However, nonexperimental approaches have estimated outcomes

for young people less well, contradicting the experimental result that adolescent relocation to low-poverty neighborhoods reduces violent crime arrests.

Similarly, nonexperimental estimates of neighborhood and peer influence on academic outcomes have differed from experimental results. While cross-sectional regressions have found little effect, randomized experiments have shown significant test-score gains. Further research on such discrepancies is needed.

The more sophisticated nonexperimental approaches have yielded important new information about education policy. The controversial Head Start program, for instance, has shown no or negative effect on children when standard models were used, but sibling-difference models, applied to the Child–Mother File of the National Longitudinal Survey, have indicated that Head Start positively affects test scores for White and Black children, though gains for Blacks fade within a few years, suggesting that it is the environment of Black children, not Head Start, that accounts for the fading of effects.

Cross-sectional and value-added analyses of another controversial issue, school funding, have yielded mixed results, and bias seems significant even with rich data sets. With IV models, however, results have been less equivocal. Additional school spending showed decisive benefits to test scores in several studies; in one study the gains were five times as large as in simpler models. Spending on class-size reductions was also shown to improve elementary student achievement in IV estimates, though not in a study based in Connecticut, where class sizes are already small. Analysis of computer spending and use yielded less conclusive results, since cross-sectional analysis showed some associated test-score improvement, but more convincing IV analysis showed no significant improvement resulting from computer use.

Whether school vouchers for private-school choice improve outcomes has also been debated, and while local voucher experiments suggest that private schools may be more effective than public, national analysis requires nonexperimental approaches. Value-added studies of NELS data have found only small test-score advantages to private schools, but self-selection bias is hard to discount here. It has been challenging to find IVs to eliminate this bias. One IV study used exogenous variation in public-transportation availability to examine nonacademic outcomes; this analysis associated private religious schooling more strongly with lower adolescent conduct problems than in simpler estimates. In sum, more sophisticated analytical models tend to show more definite results for policy interventions.

Conclusions

Review of nonexperimental policy evaluation highlights the value of sophisticated methods while emphasizing the need to scrutinize the role of self-selection in both program choice and outcomes. Analytical methods controlling for self-selection and other unobserved factors have yielded *larger* effects than less controlled models. This suggests that negative selection needs attention. That is, compensatory targeting of many programs toward at-risk children leads to understatement of program productivity in standard models, since targeting leaves many student and family variables unmeasured.

Finally, since econometric advances have improved estimations of educational program effects, researchers should use cost-effectiveness analyses more vigorously to identify programs producing the greatest student gain for given expenditure. That would help policymakers compare the effectiveness of education programs to that of housing or health interventions to help children.



(**Making**, continued from p. 1)

offered in five general areas: accountability; markets; theory; research; and working relations among analysts, policymakers, and educators.

ACCOUNTABILITY

It was agreed that accountability should be improved through better informing taxpayers and other concerned community members about how funds are distributed for achieving the school's outcomes and about how and whether outcomes are being attained. New forms of accountability and estimation—including school-site budgeting and decision making as well as longer term, value-added outcome measures—should be introduced, if carefully planned and executed.

MARKETS

A greater market role in resource allocation was recommended to improve accountability and productivity. The introduction of market mechanisms through vouchers or privately governed charter schools can motivate schools to become more efficient. Moreover, introducing market forces into teacher hiring and compensation practices, for instance by flexibly matching teacher preferences with school needs, may lead to more effective teaching. The strength of traditional practices tying compensation strictly to education and experience argues for carefully planned introduction of such changes.

THEORY

Economic theory, conferees suggested, can play a large role in improving school productivity, both through assessing productivity and enhancing our understanding of complicated educational finance issues like adequacy and equity. However, economists must consider theory and findings from other social sciences to encompass the full range of causes, effects, costs, and benefits of education policies and practices. Especially important is recognizing conflicting interests among groups involved in education decisions

(including economists). Such differences cannot be expressed in economic terms alone. It is also advisable that productivity researchers consider psychometric assessment of achievement and socio-economic effects of student outcomes.

RESEARCH

It was also suggested that economists need better data than has been generally available. They should consider using the most detailed data, at the individual-student level, since it is likely to be most valuable in precisely estimating the influence of economic factors on student outcomes. Long-term data, especially on adult outcomes, should be compiled. Mixed and complementary research methods should be used. Typical statistical analyses with large, single-point data sets should be combined with such methods as randomized experiments and small-scale, closely monitored studies of program implementation. Such intensive research may allow better understanding of what works best for specific groups of students.

WORKING RELATIONS

Participants urged future work to address pressing questions regarding the improvement of relations between those working at various levels of education service. Which government organizational strategies will improve schools most? How should state certification and professional development practices be reformed to help teachers develop more effective classroom techniques? How can private and public schools be encouraged to compete productively? How should private and public funding be integrated for optimal outcomes? What large- and small-scale reforms will work best together to ensure student success?

Conclusion

In light of these difficult, open questions and the broad areas for future research, policy, and practice recommended, the conferees agreed that further conferences like this one and dissemination of productivity findings for a wider audience are much

needed. The participants celebrated this collaborative effort to analyze and discuss solutions to problems related to continuing limitations in knowledge of what makes schools work. Through such exchange of viewpoints and finding of common ground, economists and educators can productively work together to improve students' futures.

In Memoriam: Margaret Wang

Margaret Wang provided the energy and vision that made this conference possible. The conference occurred just after her untimely and unexpected passing before Thanksgiving of 2000. Her spirit was clearly present throughout the conference and continues to guide this work. She was excited about the conference and the response it elicited. It is fitting that Margaret be connected to this theme. She understood and appreciated the importance of economics and was eager to bring this perspective more to the foreground of debate. Inspired by her memory, her colleagues welcome and encourage the further development of this work.



The CEIC *REVIEW*

Mark Rohland
Editor

This publication is supported in part by the Laboratory for Student Success at the Temple University Center for Research in Human Development and Education (CRHDE). The opinions expressed do not necessarily reflect the position of the supporting agencies, and no official endorsement should be inferred.

For more information, write CEIC, Temple University Center for Research in Human Development and Education, 1301 Cecil B. Moore Avenue, Philadelphia, PA 19122-6091, or call (215) 204-3000. Selected CEIC publications and abstracts of articles, including back issues of the CEIC Review, are available at <http://www.temple.edu/LSS/ceicrevlist.htm>

Find the papers commissioned for this conference in a book now available from Information Age Publishing:

Improving Educational Productivity



edited by
David H. Monk
Herbert J. Walberg
Margaret C. Wang

A VOLUME IN: RESEARCH IN
EDUCATIONAL PRODUCTIVITY





Temple University Center for Research
in Human Development and Education
1301 Cecil B. Moore Avenue
Philadelphia, PA 19122-6091

Non-Profit
U.S. Postage Paid
Philadelphia, PA
Permit No. 1044